EFFECT OF SAMPLE SIZES ON THE SIGNIFICANCE (RESULTS) OF PARAMETRIC STUDIES

BY DR CLIFFORD .M.AMINI

REGIONAL TRAINING AND RESEARCH INSTITUTE FOR OPEN AND DISTANCE LEARNING

(RETRIDAL) ,NATIONAL OPEN UNIVERSITY OF NIGERIA(NOUN)

VICTORIA ISLAND ,LAGOS.

Abstracts

Responses from 700 teachers on the Principals Administrative Behaviour and Teachers Job Satisfaction Questionnaire were analysed using Pearson (r) and Student's t-two parametric instruments The observed values of r and t ,using different sample sizes(50,100,200,and 300),were tested using transformed Z and F-ratio to see if the variations in the sizes of the sample has any effect on the significance of these tests.It was concluded that variations in the sizes of sample(treatment) were responsible for the variations in the significance(results) of the parametric tests.The two null hypotheses were thus rejected.

Introduction

The problem of what number of subjects will constitute a sample or will be appropriate for a sample from a given population of study has been an age-old problem among undergraduate and graduate students in the universities. This problem is not also alien to some scholars in their out-of-school research experiences .For example students would always ask their supervisors how many subjects would constitute a reasonable sample for their study or what percentage of the study population will be adequate for their researches.

The other seemingly important problem among students is which sampling techniques can be used to select an appropriate and fairly representative sample of a population .This is because there are many types of random sampling techniques that are available and can be used for the purpose of selecting a sample for a study.

These problems are important because researches in psychology, social sciences, education, and, distance learning, etc have largely been carried out within large population of subjects. Oftentimes the entire population is usually difficult to study due to limited time and finances, coupled with logistics. In such situations the researchers are left with no option than the use of a part or portion (usually a proportion) of the population that is taken for study. Maxwell (1983),defined the sample of a population as a set of data ,which is a part of the entire collection of values of interest to the researcher. A sample, according to Minnium, King, and Bear(1993), is a part of the population. McCall(1980),recognized that the sample is a subset of a population(seen as a complete set of subjects events or scores that have common characteristics ,usually referred to as parameters).

Samples are generated from populations, using different methods. Besides the techniques used in selecting a sample from a population for the purpose of a research and for generalization of population parameters, the size(s) of the sample(s) is (are) also of significance in considering the results obtained from such inferential studies. It is generally believed that the results obtained from data analysis are affected by the sizes of the used. Needless to say that these results help inform and direct the decision reached either to retain or reject the hypotheses that guide the study. This presupposes that sample sizes can affect the decision taken and can also influence the type of errors of measurement made by researchers or otherwise.

A sample used for a study must necessarily be a true or near-true representation of the population under investigation to make room for generalization. The total influences of sample sizes help to determine the size or value of significance used in decision-making in statistical analysis of research data. Thus the value of significance becomes important factor in all research decisions. It is the result of the research.

Most researchers, according to Akpan(1993),believe that statistical significant result means valid research findings and vice versa. Most researchers are unaware that statistical significance is a minor quality of statistical certainty which is directly dependent upon the size of 'n'(Cohen and Hyman,1979). This misconception of the place of significant testing reached an alarming dimension so much so that some researchers have questioned its usefulness while others have called for the abrogation of its application in education and social sciences. Carver, (1978), criticized the practice when he observed:

> *'statistical significant testing has involved more fantasy than fact. The emphasis on statistical significance over scientific significance in educational research presents a corrupt form of the scientific method; educational research would be better off if it stopped testing its results on statistical significance (p.378)'*

There are several other condemnations of statistical significant testing in education and social sciences,(Hays,1963). The problem of the present study is to highlight the influence of size of sample on the outcome of parametric studies. Parametric studies being researches in which statistical techniques are described on the bases of values of certain statistics(sample mean ,sample standard deviation etc.),which are used to make certain assumptions about parameters-population mean, population standard deviation, sd.

Methodology

The research which adopted the survey design was conducted, using seven hundred (700) teachers from the Rivers State secondary school system. The instrument- Principals 'Administrative Behaviour and Teachers' Job Satisfaction Questionnaire was administered to the 700 teachers sampled from a population of thirteen thousand six hundred and fifty three(13,653). The questionnaire, which is a 27-item instrument ,was built on a 6-point Likert format. Structured responses ranged from very strongy agreed(VSA), strongly agreed(SA), agreed(A), disagreed(DA), strongly disagreed(SD), and very strongly disagree(VSD).

Two null hypotheses were formulated to guide the study:

HO1: There is no significant influence of sample size on the Significance (results) of Pearson r.

HO2: There is no significant influence of sample size on the Significance (results) of t-test.

Data Analysis.

The data generated from the research are analysed using Pearson r,t-test,and confirmed by Fisher's Z and F-ratio test. The results are presented in the following tables A-D.

Table A: Pearson (r) for principles Administrative Behavior and Teachers Job Satisfaction at different Sample Sizes.

Size	[x	[x^2	[y	[y^2	[xy	r-obs.	r-Critical	
							0.01	
0.02	0.05							
300	592	40749	439	28719	31750	0.48	0.25	0.23	0.05
200	393	26955	293	19042	21166	0.72	0.25	0.23	0.19
100	199	13795	146	9677	1058	0.71	0.27	0.23	0.21
50	100	6889	80	4830	531	0.36	0.35	0.32	0.27

Table B. Fisher's Z^1 transformation for r-values obtained in table A.

Sizes	r		Transformed r	Z^1	Weights(N-3) Weighted Z
300	0.48	.523*	297	155.3	
200	0.72	.908*	197	178.9	
100	0.71	.887*	97	86.04	0.595*
50	0.36	.377*	47	17.7	
650			638	437.49	

Table B shows that the values of significance at various sizes were influenced by the sizes of sample (*) the null hypothesis is rejected.

Table C: t-test analysis for teachers Job Satisfaction Vs Principals Administrative Behavior by sample sizes.

Sizes	Means[$\bar{X}$=Job satisfactn,	$\bar{Y}$=administrtn]	Standard deviation		t		t-critacal
			Sd_x	Sd_y			
50	69.92	57.89	15.4	11.5	4.40	2.7	2.4 2.00
100	66	48	30.2	20.8	4.90	2.6	2.3 2.00
200	70	55	50	40.3	3.20	2.6	2.4 2.00
300	65	56	40.6	33.8	2.95	2.6	2.3 2.00

Table D: F-test analysis of sample variance for table C

Size F*	Variance Estimates			t	F-ratio
	S_1^2	S_2^2			
50	237.16	132	4.40	1.79	1.61
100	912.04	432.64	4.90	2.11	1.42
200	2500	624.09	3.20	1.54	1.39
300	1648.36	1142.44	2.95	1.44	1.34

Table D shows the F – Values are significant thus rejecting the null hypothesis.

Discussions of Findings

Two null hypotheses were formulated to guide the study, which attempts to answer the question: Are the results of significance of parametric tests influenced by variations in the sizes of samples selected for the purpose of research? These are :1There is no significant influence of sample size on the significance of Pearson r, and 2: There is no significant influence of sample size on the significance of t-test. These two tests were used to analyse data obtained from the Principals' Administrative Behavior and Teachers Job Satisfaction questionnaire. The values of 'r' observed in table A were subjected to Fisher's Z transformation analysis to test hypothesis one. The pooled (transformed) rs of 0.595,shows that the values of significance at various levels were influenced by size of sample(compare r-calculated values of 0.48,0.72,0.72,and 0.36 at various sample sizes). The null hypothesis is rejected for Pearson 'r'

The same observation is made when the t-values obtained for the t-test were subjected to an F-ratio analysis. F-values of 1.79,2.11,1.54,and 1.44 confirmed that the t-values of 4.40,4.90,3.20,and 2.95 were due to the influence of variations in sample sizes.(Compare F*of 1.61,1.42,1.34 and 1.34).

The fore-going presupposes that hypotheses one and two are rejected. This means that the values of rs and ts can react to variations in sample sizes. The results posted above agrees with the nature of correlation itself. The coefficient of correlation is a function of sampling factors. Nunary,(1960),and McCall(1980),explained the factors to be cases of (i)restriction range, (ii)extreme groups,(iii) assumptions.

Correlation in general and product moment (pm) correlation in particular is a function of the variations of the dependent variable (in this case the values of significance).Thus if a broader range of subjects is studied, the correlation will increase (as in the present study); if a narrower range is studied, the correlation is reduced. Observed correlation values at the sample size of 50 in the present study is lower than all the other sample sizes of 100, 200,and 300 respectively.(Table A ,r-obs).The critical values at 0.01, 0.02, 0.05 levels resonate around 0.19, 0.23, 0.35. This is because of the restricted range of degrees of freedom (df). Above 100, the df for pm correlation is considered infinity and yield the same scores. Thus Nunary (1967) asserts that the range in variance must be a real change arising from difference in sampling methods or sizes.

Correlation is not affected by artificial changes in variance. Thus the reactions observed in the present study for hypothesis one is considered real and due to treatment. This result agrees with Carver (1978) who averred that 'trivial' results are often interpreted as significant or important, when they are simply results that would happen when using random samples from the same population using large sample sizes-that is, a mean difference that is small and not significant from a research standpoint can be statistically significant just because enough(large sample) subjects were used in the experiment to make the result statistically rare under null hypothesis.

For t-test analysis, several results agrees with the findings of this study. For example Crombach and Snow (1977) observed that inadequate sample size could hamper the detection of interaction effect during data analysis. Also in his research article on statistical power magnitude effects and research sample size, Freedman (1982), was concerned about the misconception of the effect of sample size on statistical significant test. He explained that the basic question a researcher should consider is 'what sample size should he employ'? In a related study by Cheung and Lai,(1993),titled: 'Finite-Sample Sizes of Johansen's Likelihood Ratio test for Co-integration' the authors looked at the critical values (significant values) in a media-multivariate normality test as these relate to sample size to a very finite value. The result shows that the significant values obtained are sample size-dependent for t-test where

$$a = \frac{\dot{X}_1 - \dot{X}_2}{\sqrt{\frac{(n_1 - 1)S^2_1 + (n_2 - 1)S^2_2}{n_1 + n_2 - 2}}}$$

and a is the difference in means between groups divided by the pooled standard deviations (Hogarthy&Kromrey,1999).

The result of hypothesis two also agrees with that of Robey and Barcikowski (1992), who studied the type 1 errors probability of the student's t-test, which they said is spuriously elevated or depressed by unequal variance combined with unequal sample

sizes. Robey and Barcikowski (1992) established that the hypothesis for both significance tests are dependent on sample variance which in turn are dependent on sample sizes.

Conclusion

Based on the result revealed by the statistical analysis of data collected for the research it was concluded that the observed and critical values obtained for parametric tests (Pearson r and t-test) are influenced by the sizes of the samples used in the study. On the basis of this conclusion the following recommendations are made:

-use of significance testing in education as a basis for decision making should be de-emphasized rather approaches that are more descriptive should be used.

-students should be properly guided by their supervisors as to what percentage of their target population could constitute reasonable and representative sample of the population. 10% is recommended where the population is large; a reduced percentage can be used where the population is small.

The above recommendations are made cognisanze of the fact that the decisions that are taken daily concerning policies in education social sciences and psychology etc. are hinged on the outcomes of research findings and the conclusions that are based on them. This means that if research sample sizes are not properly controlled to the effect that faulty research conclusions are made, it follows that decisions and policies that are anchored on these findings will be faulty since they are based on faulty foundations. This will have serious implication for Education being a very important social service. It is the opinion of this study that research conclusions should be properly guided to the extent that such factors as size of sample which can impact on the significance of research should be given proper consideration.

References

Akpan, S. M. (1993).Concept and practice in statistical significance testing in education and social sciences .PH.D. Dissertation. Graduate school university of Calabar.

Carver, R. P. (1978). The case against statistical significance testing. Harvard Education Review. 48 (3),378-399.

Cheung, Y. W, & Lai, K. S. (1993). Finite-sample sizes of Johansen's likelihood ratio test for co-integration. Oxford Bulletin of Economics and Statistics.55. 313-328

Cohen, S.A & Hayman, J.S. (1979) How come so many hypothesis in Educational research are supported? (A model Proposal). Educational Researcher 8(14), 12-16.

Crombach, L.J and Snow, R.E. (1977). Aptitudes and instructional methods: A handbook for research on interactions. N. Y. Irvington.

Freedman, H. (1982). Simplified determination of statistical power magnitude effect and research sample size. Educational and psychological Journal, 42, 521-527

Hays, W.L (1963). Statistics. New york: hott, rinchart, and wiston.

Hogarthy, K.Y & Kromrey, J.D (1991). Traditional Versus robust effect size estimates: an empirical comparison in meta-analysis tests of homogeneity. Paper presented at the Annual meeting of American Educational Research Association, Montreal.

Maxwell, E. A (1983). Introduction to statistical thinking. New Jessy, Present Hall Inc.

Mc call, R.B. (1980). Fundamental Statistics for psychology (3rd Edition) N.Y. Harcourt Bruce

Minium, E.W, King, B.M & Bear, G. (1993). Statistical Reasoning in psychology and Education. N.Y. John wiley & sons inc.

Nunally, J.C (1960) the place of statistics in psychology. N.Y. McGraw – Hill

Robey, R.R & Barkcikowski, R.S (1992) Type 1 error and the number of iterations in monte carol studies of robustness. British Journal of Mathematical and statistical psychology. 45.